

**University of Groningen**

## **Kunnen gesloten toetsen bijdragen aan de toetsing van schrijfvaardigheid?**

de Glopper, Cornelis; Willemsen, Annerose

*Published in:*  
Levende Talen Tijdschrift

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Final author's version (accepted by publisher, after peer review)

*Publication date:*  
2014

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

de Glopper, C., & Willemsen, A. (2014). Kunnen gesloten toetsen bijdragen aan de toetsing van schrijfvaardigheid? *Levende Talen Tijdschrift*, 15(1), 31-38.

### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## Kunnen gesloten toetsen bijdragen aan de toetsing van schrijfvaardigheid?

Kees de Glopper & Annerose Willemsen

*Levende Talen Tijdschrift 2014-2 bevat een artikel van Theo Pullens, Hanny den Ouden, Wolfgang Herrlitz en Huub van den Bergh waarin zij de vraag stellen of een meerkeuzetoets kan bijdragen aan het meten van schriftelijke taalvaardigheid. Op basis van hun onderzoek concluderen zij dat er vraagtekens moeten worden gezet bij het gebruik van meerkeuzetoetsen als meetinstrument van schriftelijke taalvaardigheid. Zij formuleren als advies om deze toetsvorm niet langer op te nemen in de eindtoets basisonderwijs en geen plaats te geven in de te ontwikkelen diagnostische tussentijdse toets voor het voortgezet onderwijs. Het advies om meerkeuzetoetsen voor schrijfvaardigheid in de ban te doen is in onze ogen overhaast, zeker omdat het advies enkel op de eigen bevindingen van Pullens en collega's lijkt te berusten.*

In dit artikel betogen wij dat gesloten toetsen een kans verdienen in de ontwikkeling van diagnostische toetsen voor schrijfvaardigheid. Wij hopen met ons pleidooi niet alleen Theo Pullens en zijn collega's op andere gedachten te brengen. Onze boodschap is ook bestemd voor het bestuur van de Sectie Nederlands van de Vereniging van Leraren in Levende Talen. Bij monde van zijn voorzitter, Klaas Heemskerk, liet het Sectiebestuur onder verwijzing naar het onderzoek van Pullens et al. (2013) in de zomer van 2013 aan de onderwijswoordvoerders in de Tweede Kamer weten dat meerkeuzetoetsen ongeschikt zijn gebleken voor het meten van schrijfvaardigheid.

### 'Schrijfvaardigheid' en 'validiteit'

Pullens et al. (2013) rapporteren over gegevens die afkomstig zijn uit een onderzoek naar de effecten van het computerprogramma TiO-schrijven. Bij 168 leerlingen uit groep 8 van het basisonderwijs werden 12 teksten verzameld en vond een afname plaats van een van de twee parallelle subtests uit de door Van Schooten (1988) ontwikkelde meerkeuzetoets. De gegevens van de voormeting aan het begin van het schooljaar (vier opstellen en de A-variant van de toets) en de nameting aan het einde van het jaar (opnieuw vier opstellen en de A-variant) werden door Pullens et al. geanalyseerd. Zij constateren dat de meerkeuzetoets geen voorspellende waarde heeft voor oordelen over de kwaliteit van de teksten. Alleen de deelttoets Spelling & Interpunctie hangt statistisch significant samen met zogenoemde 'schaaloordelen' over schrijfvaardigheid. Deze schaaloordelen zijn per leerling voor de voor- en de nameting gebaseerd op alle vier de teksten, waarbij iedere tekst is beoordeeld in vergelijking met een opdrachtspecifieke voorbeeldtekst van gemiddelde kwaliteit. De drie deelttoetsen voor Inhoud & Organisatie, Idioom & Stijl en Grammatica dragen niet bij aan de voorspelling van de schaaloordelen. Dit wijst volgens de onderzoekers op een lage constructvaliditeit van de meerkeuzetoets. Deze meet, in alledaagse woorden, niet wat hij meten moet: het construct of begrip 'schrijfvaardigheid'. Dat ligt volgens Pullens et al. niet aan eventuele tekortkomingen van de meerkeuzetoets: de kwaliteit van de meerkeuze-items en de interne structuur van de toets als geheel lijken in orde en met drie van de vier deelttoetsen (Idioom & Stijl, Grammatica en Spelling & Interpunctie) kunnen verschillen tussen de voor- en nameting worden geregistreerd.

De kwestie die Pullens en zijn collega's aansnijden betreft dus de validiteit van instrumenten voor het meten van schrijfvaardigheid. Wat bedoelen we als we het over de validiteit of geldigheid van een meetinstrument hebben? Met Borsboom & Mellenbergh (2007) verstaan we daar dit onder: "A test is valid for measuring a theoretical attribute if and only if variation in the attribute causes variation in the measurement outcomes through the response processes that the test elicits" (p.93).

Vertaald naar ons domein betekent dit het volgende. Schrijfvaardigheid is een theoretisch attribuut: we kunnen het niet observeren, maar wel aan personen toeschrijven, *in de regel* op grond van verschillen in de kwaliteit van de teksten die zij schrijven. Bij een valide test veroorzaken verschillen in schrijfvaardigheid verschillen in toetsscores. De scoreverschillen komen daarbij tot stand door de processen die worden opgeroepen bij de uitvoering van de schrijfo opdracht. De validiteitskwestie komt daarom hier op neer: als het goed is worden variaties in schrijfvaardigheid door de antwoordprocessen bij het afleggen van de toets overgedragen naar variaties in scores.

Het maakt daarbij op zichzelf niet uit of de toets open of gesloten is. Bij een open schrijftoets moet een schrijfo opdracht worden uitgevoerd en moet de tekst op zijn kwaliteit worden beoordeeld. Het is bij open toetsen daarom de vraag of verschillen in schrijfvaardigheid zich tijdens de constructie van de tekst manifesteren én of ze tot uitdrukking komen in de beoordeling van de kwaliteit van de tekst. De schrijftoets kan ook een meer gesloten vorm aannemen, zoals de correctieopdracht die verlangt dat de schrijver tekortkomingen in een gegeven tekst opheft, of de meerkeuzetoets waarbij de schrijver voor een reeks van opgaven steeds de juiste verbetering of het beste formuleringsalternatief kiest. Ook bij geheel gesloten toetsen is het de vraag of bij het maken van de opgaven de schrijfvaardigheid wordt aangesproken.

Van het beantwoorden van de vraag naar de validiteit van metingen van schrijfvaardigheid zouden wij ons kunnen afmaken door de tweede zin van de alinea voor de vorige alinea in verkorte en licht gewijzigde vorm te herhalen: schrijfvaardigheid is een eigenschap die wij aan personen toeschrijven op grond van verschillen in de kwaliteit van de teksten die zij schrijven. Zo gesteld kunnen alleen open opdrachten valide metingen van schrijfvaardigheid opleveren. Toetsprocedures waarbij schrijvers niet zelf tekst moeten produceren, zijn dan per definitie invalide. Ons lijkt een ietwat ruimere formulering van het begrip schrijfvaardigheid verstandiger: *schrijfvaardigheid is een eigenschap die wij aan personen toeschrijven op grond van verschillen in de kwaliteit waarmee zij problemen in de schriftelijke communicatie oplossen*. De winst van deze formulering zit hem wat ons betreft in het gebruik van het begrip ‘(schrijf)probleem’. Met dit begrip denken wij de kenmerken en de voor- en nadelen van de verschillende meetmethoden scherper te kunnen stellen. In de volgende paragraaf doen wij daar een poging toe.

### **Schrijfproblemen in open en gesloten opdrachten**

We gebruiken het begrip schrijfprobleem hier in de volgende betekenis: *van een schrijfprobleem is sprake wanneer er een discrepantie bestaat tussen een wenselijke en een feitelijke toestand in de schriftelijke communicatie en het niet duidelijk is hoe het doel moet worden bereikt*. Met deze formulering sluiten we aan bij de klassieke en in de psychologie breed aanvaarde formulering (Duncker, 1945; Mayer, 2013) van cognitieve (en sociaalcognitieve) problemen.

Een open schrijfo opdracht stelt een *complex* probleem: om een wenselijke toestand te creëren (die bijvoorbeeld kan bestaan uit gedeelde kennis, een geslaagde handeling of een gedeelde opvatting) moet een tekst worden geschreven waarmee bepaalde handelingen worden verricht (zoals informeren, instrueren, overtuigen). Het gestelde probleem is complex omdat er problemen op verschillende niveaus moeten worden opgelost: op het vlak van de inhoud, samenhang, opbouw, formulering, spelling en interpunctie moet werk worden geleverd. Werk waaraan schrijvers hun handen vaak meer dan vol hebben.

Open schrijfo opdrachten stellen problemen die niet alleen complex maar ook *diffuus* zijn. Zij bieden schrijvers veel ruimte: ruimte voor het interpreteren van het in de opdracht gestelde communicatieve probleem, ruimte

voor het kiezen uit alternatieve oplossingen ervan en ruimte voor het vermijden van lokale problemen in bijvoorbeeld de samenhang, formulering of spelling.

De problemen die open opdrachten stellen zijn verder eenmalig of *uniek*. Open opdrachten worden daarom ook wel als één-item-toetsen aangemerkt (De Glopper, 1989; Schoonen, 1991). Een open opdracht biedt de schrijver één kans op het tonen van zijn vaardigheid-als-geheel. Voor het tonen van 'deelvaardigheden' hangt het aantal kansen af van de aard van het deelprobleem in kwestie. Gaat het om het formuleren van grammaticaal correcte zinnen of om het aanbrengen van variatie in zinsopeningen? Dan kan bij een wat langere tekst het aantal kansen best flink zijn. Maar is bijvoorbeeld de keuze voor en de correcte toepassing van een bepaald argumentatieschema in het geding, dan is het voor de schrijver in één keer erop of eronder.

Bij gesloten schrijfp opdrachten zoals de meerkeuzetoets worden *enkelvoudige* problemen gesteld. Het kan daarbij om allerlei kwesties gaan, zoals de selectie van relevante inhoud voor een bepaalde tekst of om de keuze voor een gepaste formulering, gelet op de context en het lezerspubliek. De problemen in meerkeuzetoetsen zijn ook *gefocust*: de vrijheid voor de schrijver is ingeperkt, want zij bieden geen ruimte voor interpretatie, eigen keuzes of vermindering. Ten slotte bevatten gesloten toetsen *herhaalde instanties* van problemen: zij bieden schrijvers meerdere kansen voor het oplossen van een zelfde type probleem.

### **Bezwaren tegen de validiteit van gesloten toetsen voor schrijfvaardigheid**

Tegen gesloten toetsen voor schrijfvaardigheid wordt als voornaamste bezwaar ingebracht dat de schrijver geen eigen tekst moet produceren, maar in een gegeven tekst gegeven problemen moet oplossen. Tegen dit evidente en tenminste op het eerste gezicht ernstige bezwaar kan het volgende worden ingebracht.

Schrijven doet een beroep op allerlei vormen van kennis. Schrijvers moeten gebruik maken van kennis van het publiek waarvoor en het onderwerp waarover zij schrijven, van het te beoefenen genre, van de te gebruiken taal en van de conventies van het schrift (Byrnes, 2009). Deze vormen van kennis kunnen ook worden aangeboord met gesloten toetsen. Voor het oplossen van gegeven tekstproblemen moeten schrijvers immers ook gebruik maken van kennis. Dat gesloten toetsen schrijvers daarbij laten lezen in plaats van schrijven is waar, maar reflectief of kritisch lezen is wel een activiteit die volgens recente theorieën een centrale plaats inneemt in het schrijfproces (Hayes, 1996).

Een ander bezwaar tegen gesloten toetsen is het volgende. Door de focus op steeds één probleem tegelijk verlangen gesloten toetsen niet wat zo kenmerkend is voor schrijven: 'juggling of constraints' (Flower & Hayes, 1980) en orkestratie van activiteiten (Graham & Harris, 2000). Terwijl schrijvers bij de productie van eigen tekst hun aandacht moeten verdelen over schrijfproblemen op verschillende niveaus, kunnen zij zich bij het afleggen van een gesloten toets concentreren op steeds één probleem tegelijk. Deze onmiskenbare reductie brengt echter ook een voordeel met zich mee. Bij open opdrachten kunnen schrijfproducten eenzelfde kwaliteitsscore krijgen terwijl de teksten en de betreffende schrijfprocessen door allerlei verschillende sterke en zwakke punten gekenmerkt kunnen worden. Bij gesloten toetsen blijven de verschillende zwakke en sterke punten gescheiden en daardoor zichtbaar.

Open toetsen hebben in vergelijking met gesloten toetsen één ijzersterke charme: hun indrukvaliditeit. Iedereen ziet immers onmiddellijk in dat zij schrijvers op de proef stellen door ze laten doen waar het om draait: teksten schrijven. Open toetsen voor schrijfvaardigheid hebben daarom wat validiteit betreft onmiskenbaar een enorme potentie. Of die potentie ook wordt gerealiseerd is echter maar de vraag.

We moeten ten eerste wijzen op de problematiek van het beoordelen van de teksten die bij open toetsvormen worden geschreven. De welbekende problemen die daarbij spelen (De Groot, 1961; Wesdorp, 1981) zijn voornamelijk<sup>1</sup> onlosmakelijk verbonden aan de open meetmethoden. Bij inzet van voldoende middelen, deskundigheid en bereidheid tot het delen van criteria en standaarden zijn beoordelingsproblemen tot op grote hoogte oplosbaar, maar in de dagelijkse praktijk van onderwijs en toetsing zijn deze voorwaarden niet altijd vervuld. Dit maakt dat de reële bestaande validiteit van open toetsen achterblijft bij hun validiteitspotentie. We willen hier ten tweede onze karakterisering herhalen van de problemen die open toetsen stellen: deze problemen zijn complex, diffuus en uniek. Ook daardoor kan afbreuk worden gedaan aan de validiteit van open metingen. Bezwaren tegen de validiteit komen dus niet alleen aan gesloten toetsen toe.

### **Empirische evidentie voor de validiteit van gesloten toetsen voor schrijfvaardigheid**

Tegen de bovenstaande theoretische beschouwingen kan worden ingebracht dat de status ervan op zijn best voorlopig is. Hoe het werkelijk is gesteld met de validiteit van toetsen voor schrijfvaardigheid zal toch uit empirisch onderzoek moeten blijken. Wat *weten* we inmiddels, bijna een halve eeuw na de publicatie van de klassieke studie van Godschalk, Swineford en Coffman (1966)? We beantwoorden deze vraag voor gesloten toetsing: de toetsvorm die in de beklagenbank zit. Aan ons antwoord gaan drie waarschuwingen vooraf.

Waarschuwing één: afzonderlijke studies naar de validiteit van metingen van schrijfvaardigheid zijn geen van alle doorslaggevend, omdat ieder onderzoek zijn onvermijdelijke beperkingen kent. De uitkomsten van het validiteitsonderzoek zijn mede daarom sterk variabel, zoals duidelijk wordt uit de overzichten die Van Schooten & De Gloppe (1990) en Schoonen (1991) bieden. Het is in dit licht niet verstandig om op basis van de uitkomsten van een enkele studie te 'generaliseren naar een lage constructvaliditeit van meerkeuzetoetsen in het algemeen' (Pullens et al., 2013, p. 40), zeker niet wanneer het een studie betreft waarvan de uitkomsten sterk atypisch lijken. Dat meerkeuzetoetsen geen voorspellende waarde hebben voor open metingen van schrijfvaardigheid is namelijk tamelijk uitzonderlijk. In de empirische studies die volgens Van Schooten en De Gloppe (1990) een adequate opzet kennen, bedraagt de mediane correlatie tussen meerkeuzetoets en open toets .82. Dit betekent dat het heel gewoon is dat twee derde deel van de variatie in de scores op open schrijfofdrachten voorspeld kan worden op basis van scores op meerkeuzetoetsen.

Waarschuwing twee: eerst binnen het kader van een theorie krijgen uitkomsten van empirisch onderzoek betekenis, ook (of misschien wel juist) wanneer er complexe statistische analyses in het geding zijn. Borsboom & Mellenbergh (2007) maken hier behartenswaardige opmerkingen over. Zij wijzen er bijvoorbeeld op dat passende statistische modellen voor de interne structuur van toetsen en hun onderdelen (hier: deelttoetsen voor bijvoorbeeld inhoud, samenhang, opbouw en taalgebruik) geen bewijs vormen dat het construct in kwestie (hier: schrijfvaardigheid) dezelfde structuur kent. Strikt genomen krijg je bij onderzoek naar de interne validiteit van een schrijftoets, zoals uitgevoerd door Van Schooten en De Gloppe (1991) en door Pullens et al. (2013) op zijn best uit de toetsing wat je erin stopt: de bedoelde factorstructuur. Hoewel dat geen geringe onderzoeksprestatie is, moet het belang ervan niet worden overschat. Een passend meetmodel voldoet niet als validatie van het te meten construct. In ongunstige gevallen is het niet meer dan een mathematisch verantwoorde, zuinige en zuivere samenvatting van scores op items en deelttoetsen.

---

<sup>1</sup> *Vooralsnog*: er zijn weliswaar belangwekkende ontwikkelingen gaande op het vlak van de automatische beoordeling van de kwaliteit van teksten (Shermis & Burstein, 2013), maar het ziet er ons inziens niet naar uit dat de computer de menselijke beoordelaar snel gaat verdringen.

Waarschuwing drie: wat we weten uit onderzoek naar de validiteit van metingen van schrijfvaardigheid heeft vrijwel zonder uitzondering betrekking op *differentiële toetsing*: het meten van individuele verschillen in schrijfvaardigheid. Het gaat daarbij om het voorspellen van de rangordering van schrijvers op één dimensie: een rechte lijn waarop zij naar vaardigheid gerangschikt kunnen worden. In de literatuur hebben wij geen onderzoek kunnen vinden naar de specifieke bijdrage van gesloten toetsen aan *diagnostische toetsing* van schrijfvaardigheid, de vorm van toetsing die aan de orde is in de brief van het Sectiebestuur Nederlands aan de onderwijswoordvoerders in de Tweede Kamer.

Wij vinden het van groot belang om op dit punt een scherp onderscheid te maken tussen beoordeling van tekstkwaliteit en diagnostische toetsing van schrijfvaardigheid. Tekstkwaliteit wordt vaak beoordeeld in situaties waarin sprake is van differentiële meting : de kwaliteit van een tekst wordt dan bepaald in vergelijking met die van andere teksten. Bij metingen van tekstkwaliteit gaat het geregeld ook om het vaststellen van de mate waarin een de tekst het (door de opdracht opgelegde) doel van de schrijver dient: de informatiewaarde, overtuigingskracht of behulpzaamheid van de tekst wordt dan vastgesteld in relatie tot een concreet informatief, persuasief of instructief doel. Bij diagnostische toetsing wordt aan de hand van teksten van leerlingen vastgesteld in hoeverre er sprake is van specifieke schrijfproblemen. Het gaat bij diagnostische toetsing niet om een vergelijking met anderen, maar om het vaststellen van de aard, de ernst en de frequentie van specifieke soorten fouten en onvolkomenheden. Onze omschrijving van diagnostische toetsing past binnen de traditie van de cognitieve diagnostiek (Leighton & Gierl, 2007). Deze benadering is er op gericht specifieke cognitieve toestanden of structuren vast te stellen die oorzakelijk verbonden zijn met manifeste problemen in de uitvoering en uitkomsten van taken. Cognitieve diagnostiek is zo'n drie decennia in ontwikkeling, op het snijvlak van cognitieve psychologie en psychometrie.

Tegen de achtergrond van deze waarschuwingen, schetsen wij nu in het kort de empirische evidentie voor de validiteit van gesloten toetsen voor schrijfvaardigheid.

Voor de meetmodellen die onder gesloten toetsen liggen is in onderzoek dat in Nederland is uitgevoerd herhaalde malen ondersteuning gevonden. Van Schooten en De Glopper (1991) constateerden dat de theoretische indeling in deelvaardigheden van de door Van Schooten (1988) geconstrueerde toets bij afname onder leerlingen uit groep 8 van het basisonderwijs door de empirie gedeeltelijk werd ondersteund. Van Schooten's toets is ook gebruikt door Schoonen (1991) en Pullens et al. (2013). Schoonen (1991) besteedt aan de interne structuur van de door hem gebruikte deelttoetsen nauwelijks aandacht. Pullens et al. (2013) laten zien dat de vier deelttoetsen die zij gebruikten in de empirie goed onderscheidbaar zijn. In een onderzoek onder universitaire studenten is Meuffels (1996) overwegend positief over de interne structuur van een door hem ontwikkelde schrijfvaardigheidstoets. Een review van in het buitenland uitgevoerde analyses van de interne structuur van gesloten toetsen voor schrijfvaardigheid kennen wij helaas niet.

Terwijl we over de bijdrage van gesloten toetsen aan diagnostische toetsing van schrijfvaardigheid goeddeels in het duister tasten, is de mate waarin gesloten toetsen voorspellend zijn voor oordelen over de kwaliteit van teksten vele malen onderzocht. Hierboven meldden wij al de sterke variabiliteit van de uitkomsten van onderzoek naar wat wel de criteriumvaliditeit van meerkeuzetoetsen wordt genoemd. Bij zulk onderzoek is de kwaliteit van het te voorspellen criterium vanzelfsprekend erg belangrijk: onbetrouwbare oordelen over de kwaliteit van teksten laten zich nu eenmaal niet goed voorspellen. Wat verder precies bepalend is voor de hoogte van de correlaties tussen meerkeuzetoetsen en opsteloordeelen wordt uit de overzichten die Van Schooten & De Glopper (1990) en Schoonen (1991) bieden helaas niet duidelijk. Het ontbreekt aan een systematische meta-analyse van het beschikbare onderzoek. Wel is uit eerder en ander onderzoek dan dat

van Pullens et al. (2013) duidelijk dat meerkeuzetoetsen *kunnen* bijdragen aan de voorspelling van kwaliteitsoordelen over teksten. We herhalen op deze plaats nog maar eens de hierboven vermelde mediane correlatie tussen meerkeuzetoetsen en open toetsen van .82.

Ook een recente, eigen empirische beproeving heeft ons nog eens doordrongen van de potentie van gesloten toetsen. Een door de tweede auteur ontwikkelde meerkeuzetoets voor samenhang en structuur hebben wij in verband gebracht met schaaloordeelen over 3 verschillende teksten van 84 leerlingen uit leerjaar 1, 2 en 3 van het voortgezet onderwijs. De gesloten toets bevatte deelttoetsen voor (i) het beoordelen van de relevantie van zinnen binnen gegeven alinea's, (ii) het beoordelen van de volgorde van zinnen binnen alinea's, (iii) het aanbrengen van alinea grenzen en (iv) het ordenen van alinea's. De homogeniteit van de deelttoetsen was zeer beperkt: de Cronbach alpha varieerde tussen .22 en .53. Deze lage waarden kunnen het gevolg zijn van het beperkte aantal opgaven per deelttoets (6) en van de heterogeniteit van de kennis en vaardigheden waarop de deelttoetsen een beroep doen. De deelttoetsen, die ieder slechts één van de vele aspecten van schrijfvaardigheid bedoelen te meten, vertonen desondanks in 10 van 12 gevallen significante correlaties met de schaaloordeelen voor de drie teksten. De score op de meerkeuzetoets als geheel voorspelt 18 tot 42% van de variatie in de 3 schaaloordeelen.

## Conclusie

De conclusie dat meerkeuzetoetsen niet kunnen bijdragen aan de differentiële meting van schrijfvaardigheid achten wij in het licht van het bovenstaande eenzijdig. Gevolgtrekkingen ten aanzien van hun bijdrage aan diagnostische toetsing zijn ons inziens overhaast, want deze missen grond in onderzoek. Bij de inzet van open schrijfvaardigheidstoetsen voor diagnostische doeleinden passen ook vraagtekens, aangezien deze toetsen schrijvers met complexe, diffuse en unieke problemen confronteren. Onze twijfel aan de diagnostische waarde van open toetsen wordt versterkt door de beoordelingsproblemen die in de praktijk van onderwijs en toetsing vigeren. Ons advies inzake de ontwikkeling van diagnostische toetsen voor schrijfvaardigheid luidt daarom anders dan dat van Pullens et al. (2013) en het Sectiebestuur Nederlands: onderzoek alles en behoud het goede.

## Literatuur

- Borsboom, D. & Mellenbergh, G.J. (2007). Test validity in cognitive assessment. In: J.P. Leighton & M.J. Gierl (Eds.). *Cognitive diagnostic assessment for education: Theory and applications* (p. 85-115). Cambridge: Cambridge University Press.
- Byrnes, J. P. (2009). *Cognitive development and learning in instructional contexts*. Boston, Mass.; London: Allyn and Bacon.
- Duncker, K. 1945. *On problem solving*. Psychological Monographs, 58, no. 5: Whole No. 270.
- Flower, L., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. *Cognitive processes in writing*, 31, 50.
- Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist*, 35, 3-12.

- Groot, A.D. de (1961). *Methodologie. Grondslagen van onderzoek en denken in de gedragswetenschappen*. Den Haag: Mouton & Co.
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (red.) *The science of writing: Theories, methods, individual differences, and applications* (p. 1-27). Hillsdale, NJ, England: Lawrence Erlbaum Associates.
- Leighton, J.P., & Gierl, M.J. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge: Cambridge University Press.
- Mayer, R.E. 2013. Problem solving. In: D. Reisberg (Ed.). *The Oxford handbook of cognitive psychology* (p. 769-778). New York: Oxford University Press.
- Meuffels, B. (1996). *Diagnostische schrijfvaardigheidstoetsen: De constructie van objectief of semi-objectief te scoren diagnostische schrijfvaardigheidstoetsen voor schrijvers van universitair niveau*. Amsterdam: Institute for Functional Research into Language and Language Use.
- Pullens, T., Den Ouden, H., Herrlitz, W., & Van den Bergh, H. (2013). Kan een meerkeuzetoets bijdragen aan het meten van schriftelijke taalvaardigheid? *Levende Talen Tijdschrift*, 14(2), 31-41.
- Schoonen, R. (1991). *De evaluatie van schrijfvaardigheidsmetingen: Een empirische studie naar betrouwbaarheid, validiteit en bruikbaarheid van schrijfvaardigheidsmetingen in de achtste groep van het basisonderwijs*. Academisch proefschrift. Amsterdam: Universiteit van Amsterdam.
- Schooten, E. van, & Glopper, K. de (1990). De validiteit van meerkeuze-instrumenten voor het meten van schrijfvaardigheid. *Tijdschrift voor Taalbeheersing*, 12, 93-110.
- Schooten, E. van & Glopper, K. de (1991). De constructie en interne validering van een meerkeuzetoets voor het meten van schrijfvaardigheid. *Tijdschrift voor Onderwijsresearch*, 16(2), 72-92.
- Shermis, M. D., & Burstein, J. (2013). *Handbook of automated essay evaluation: Current applications and new directions*. New York, NY US: Routledge/Taylor & Francis Group.
- Wesdorp, H. (1981). *Evaluatietechnieken voor het moedertaalonderwijs: Een inventarisatie van beoordelingsmethoden voor stelvaardigheid, het begrijpend lezen, de spreek-, luister- en discussievaardigheid*. Den Haag: S.V.O.